Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other
Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP
Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

How to lower the ETL barrier going to OMOP with Perseus (Demonstration)
Presenter: Zachary Wang • Graduate Student, Johns Hopkins (2022 Kheiron Cohort member)

Lowering the deployment burden with the cloud
Presenter: Lee Evans • Owner, LTS Computing LLC

Distributed Machine Learning Using OMOP
Presenter: Emily Pfaff • Research Assistant Professor, University of North Carolina at Chapel Hill
“OHDSI presents the opportunity to lower the cost of clinical registries by orders of magnitude through standardized EHR automation.”
Clinical registries have a high human cost in chart abstraction

Trauma registry methodology: A survey of trauma registry custodians to determine current approaches

Gerard M. O’Reilly, Belinda Gabbe, Peter A. Cameron

Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Commercial Rd, Melbourne, 3004, Australia

Emergency and Trauma Centre, Alfred Health, Commercial Rd, Melbourne, Victoria 3004, Australia

Emergency Services, Hamad Medical Corporation, Doha, Qatar

0.5 FTE for every 200-300 patients.

Table 3
Human resources—Single hospital registries (n=40).

<table>
<thead>
<tr>
<th>Staffing type</th>
<th>Number of persons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>All (total) staff</td>
<td>0</td>
</tr>
<tr>
<td>Director/Head</td>
<td>14</td>
</tr>
<tr>
<td>Manager</td>
<td>18</td>
</tr>
<tr>
<td>Data manager</td>
<td>17</td>
</tr>
<tr>
<td>Database programmer</td>
<td>33</td>
</tr>
<tr>
<td>Database analyst</td>
<td>31</td>
</tr>
<tr>
<td>Trauma nurse coordinator</td>
<td>17</td>
</tr>
<tr>
<td>Data collector</td>
<td>12</td>
</tr>
<tr>
<td>Data entry clerk</td>
<td>23</td>
</tr>
<tr>
<td>ICD coder</td>
<td>26</td>
</tr>
<tr>
<td>AIS coder</td>
<td>21</td>
</tr>
<tr>
<td>Data analyst</td>
<td>28</td>
</tr>
<tr>
<td>Office administrator</td>
<td>29</td>
</tr>
</tbody>
</table>

Courtesy – Jon Duke, MD, Georgia Tech Research Institute
Mapping registry data to the OMOP CDM facilitates more efficient collaborations between researchers and establishment of federated data networks.
CURE ID and Virus COVID-19 Registry

• **CURE ID**
  – Joint initiative between FDA, NIH/NCATS, Critical Path Institute
  – Aims to identify repurposed drug candidates to treat infectious diseases
  – Started as an online/app-based registry for clinicians to enter case reports
  – COVID-19 expands mandate: automated extraction from electronic health records

• **Viral Infection & Respiratory Illness Universal Study (VIRUS) COVID-19 Registry**
  – Launched by the Society of Critical Care Medicine within weeks of pandemic onset
  – Rapidly described COVID-19 clinical course
  – Global reach to 306 sites in 28 countries
  – Started with manual data entry – hundreds of variables, ~4 hours per patient
Goals for Cure ID

• Making OMOP accessible to community hospitals.
  – Create a feasibility checklist for clinical registries to adopt OMOP.
  – Lower the cost to ETL EHR data into the OMOP format.
  – Make it easier to deploy the OHDSI tool stack into a secure health system environment.
Goal 1: Clinical Registry OHDSI feasibility checklist

<table>
<thead>
<tr>
<th>Question</th>
<th>Example</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many of the data elements correspond to core OMOP concepts?</td>
<td>Meds, Labs, Procedures, Conditions, Devices</td>
<td>Out of the Box</td>
</tr>
<tr>
<td>How many of data elements will need custom transformation scripts?</td>
<td>Vitals, Epic Flowsheets, Epic SmartForms</td>
<td>Minor Effort</td>
</tr>
<tr>
<td>How many of the data elements will need a custom concept?</td>
<td>Questions that are not in any standard lexicon.</td>
<td>Minor Effort</td>
</tr>
<tr>
<td>Will the OMOP data model needs to be extended to support the registry?</td>
<td>Imaging</td>
<td>Significant Effort</td>
</tr>
<tr>
<td>How many of the data elements will require NLP?</td>
<td>Symptoms at the time of admission.</td>
<td>Significant Effort</td>
</tr>
</tbody>
</table>
Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other
Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP
Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

• Matthew Robinson, MD is an infectious disease physician at JHU who also serves as faculty in Biomedical Informatics.
Adapting a manual clinical registry to OMOP

• Evaluate the feasibility of each variable for inclusion
• Weigh the challenge versus value of variables that require significant effort
• Consider the use case from the beginning
• Identify alternative means of data representation for complex variables
• Craft reproducible methods for variables requiring custom transformation
A Challenge - WHO Scale for COVID-19

- Not present as structured data in the EMR and infrequently documented in notes
- Identification of symptoms requires NLP
- Includes variables infrequently included in an OMOP ETL
  - Respiratory support modality
    - Some modes of respiratory support lack standard terms
  - Fraction of inspired oxygen

<table>
<thead>
<tr>
<th>Patient State</th>
<th>Descriptor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninfected</td>
<td>Uninfected; no viral RNA detected</td>
<td>0</td>
</tr>
<tr>
<td>Ambulatory mild disease</td>
<td>Asymptomatic; viral RNA detected</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Symptomatic; independent</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Symptomatic; assistance needed</td>
<td>3</td>
</tr>
<tr>
<td>Hospitalised: moderate disease</td>
<td>Hospitalised: no oxygen therapy*</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Hospitalised: oxygen by mask or nasal prongs</td>
<td>5</td>
</tr>
<tr>
<td>Hospitalised: severe diseases</td>
<td>Hospitalised: oxygen by NIV or high flow</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Intubation and mechanical ventilation, $pO_2/FiO_2 &lt; 150$ or SpO$_2/FiO_2 &lt; 200$</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Mechanical ventilation $pO_2/FiO_2 &lt; 150$ (SpO$_2/FiO_2 &lt; 200$) or vasopressors</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Mechanical ventilation $pO_2/FiO_2 &lt; 150$ and vasopressors, dialysis, or ECMO</td>
<td>9</td>
</tr>
<tr>
<td>Dead</td>
<td>Dead</td>
<td>10</td>
</tr>
</tbody>
</table>

*If hospitalised for isolation only, record status as for ambulatory patient.

Figure: WHO clinical progression scale

ECMO = extracorporeal membrane oxygenation. $FiO_2$ = fraction of inspired oxygen. NIV = non-invasive ventilation. $pO_2$ = partial pressure of oxygen. $SpO_2$ = oxygen saturation.

Do we need this difficult variable – WHO Scale?

- Defines inclusion criteria and endpoints for COVID-19 trials
- Causal inference models must match similarly ill patients to isolate the impact of a therapeutic on clinical outcomes
- Lack of context from incomplete registry data contributes to misinterpretation

Patient A had an oxygen saturation of 95%

so I compared him to

Patient B had an oxygen saturation of 95%

He received Drug X and survived

Drug X saves lives!

He did not receive Drug X and died

https://www.nhlbi.nih.gov/health/ventilator
A Challenge - WHO Scale for COVID-19

- **Feasibility**
  - Out of the box to significant effort
- **Use case and needs**
  - Inpatient use requires only 4-10
  - Dropping gradations of mild disease obviates need for NLP
- **Custom transformations**
  - Respiratory support modality documented in flowsheets

<table>
<thead>
<tr>
<th>Patient State</th>
<th>Descriptor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninfected</td>
<td>Uninfected; no viral RNA detected</td>
<td>0</td>
</tr>
<tr>
<td>Ambulatory mild disease</td>
<td>Asymptomatic; viral RNA detected</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Symptomatic; independent</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Symptomatic; assistance needed</td>
<td>3</td>
</tr>
<tr>
<td>Hospitalised: moderate disease</td>
<td>Hospitalised; no oxygen therapy*</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Hospitalised; oxygen by mask or nasal prongs</td>
<td>5</td>
</tr>
<tr>
<td>Hospitalised: severe disease</td>
<td>Hospitalised; oxygen by NIV or high flow</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Intubation and mechanical ventilation, pO$_2$/FiO$_2$ &gt;150 or SpO$_2$/FiO$_2$ &gt;200</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Mechanical ventilation, pO$_2$/FiO$_2$ &lt;150 (SpO$_2$/FiO$_2$ &lt;200) or vasopressors</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Mechanical ventilation, pO$_2$/FiO$_2$ &lt;150 and vasopressors, dialysis, or ECMO</td>
<td>9</td>
</tr>
<tr>
<td>Dead</td>
<td>Dead</td>
<td>10</td>
</tr>
</tbody>
</table>

*Figure: WHO clinical progression scale
ECMO=extracorporeal membrane oxygenation. FiO$_2$=fraction of inspired oxygen. NIV=non-invasive ventilation.
pO$_2$=partial pressure of oxygen. SpO$_2$=oxygen saturation.*If hospitalised for isolation only, record status as for ambulatory patient.

Wading through the flowsheets

- Flowsheets highly customized to individual EHR deployments
- Context of flowsheet entry matters
- Find flowsheet entries by name and by content
- Source to concept map must be customized to each site
- Protocols can improve reproducibility

<table>
<thead>
<tr>
<th>Flowsheet entry name</th>
</tr>
</thead>
<tbody>
<tr>
<td>LH Resp Dev</td>
</tr>
<tr>
<td>RN Oxygen Device</td>
</tr>
<tr>
<td>O2 Delivery H321</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flowsheet entry value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room air</td>
</tr>
<tr>
<td>Ventilator</td>
</tr>
<tr>
<td>Nasal cannula</td>
</tr>
<tr>
<td>Room air</td>
</tr>
</tbody>
</table>
Goal 2: Lower the cost of ETL into the OMOP format

• Invest in the Perseus project
• An open source one stop ETL web tool
• A web-based GUI for ETL of OHDSI
  – Rabbit in the Hat
  – White Rabbit
  – Usagi
• Moved into OHDSI/Github Org on 2022-07-08
• Created a base configuration for common EHR’s
  – Epic
  – Cerner (coming)
Zachary Wang is currently a Master’s student at JHU in Biomedical Informatics and Data Science.

He is also one of our OHDSI Khierons!
Goal 3: Make it easier to deploy the OHDSI tool stack into Health Systems

• Improves data quality (DQD/Achilles)
• Helps Health System get greater value out of OHDSI
• Cure ID invested into the OHDSI BroadSea project
• Hosted with Microsoft Azure
• Docker Containers
• Security compliance
Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other
Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP
Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

How to lower the ETL barrier going to OMOP with Perseus (Demonstration)
Presenter: Zachary Wang • Graduate Student, Johns Hopkins (2022 Kheiron Cohort member)

Lowering the deployment burden with the cloud
Presenter: Lee Evans • Owner, LTS Computing LLC
# Solution Components Overview

<table>
<thead>
<tr>
<th>Database Conversion &amp; Observational Research Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perseus, Atlas, RStudio + HADES</td>
</tr>
<tr>
<td>OHDSI Broadsea Containers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Azure Cloud Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Networking, Application Servers</td>
</tr>
<tr>
<td>App Database &amp; OMOP CDM Database</td>
</tr>
</tbody>
</table>
Perseus Open Source  Extract Transform & Load (ETL)

01. Scan Native Database
   - White Rabbit

02. Load base system mapping specification & customize
   - Rabbit In A Hat

03. Convert Data to OMOP CDM

04. Map local codes to OMOP standard concepts
   - Usagi

05. Data Quality Assurance
   - Data Quality Dashboard

Sergey Samus & his team
OHDSI HADES Open Source observational analytics R packages

Population-level estimation
- **CohortMethod**
  - New-user cohort studies using large-scale regression for propensity and outcome models.
  - Learn more…
- **SelfControlledCaseSeries**
  - Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.
  - Learn more…
- **SelfControlledCohort**
  - A self-controlled cohort design, where time preceding exposure is used as control.
  - Learn more…
- **Capr**
  - Develop and manipulate complex cohort definitions in R
  - Learn more…
- **CirceR**
  - An R wrapper for Circe, a library for creating cohort definitions, expressing them as JSON, SQL, or Markdown.
  - Learn more…
- **CohortGenerator**
  - Instantiating cohorts in a database based on a set of cohort definitions.
  - Learn more…

Patient-level prediction
- **EvidenceSynthesis**
  - Routines for combining causal effect estimates and study diagnostics across multiple data sites in a distributed study.
  - Learn more…
- **PatientLevelPrediction**
  - Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.
  - Learn more…
- **EnsemblePatientLevelPrediction**
  - Building and validating ensemble patient-level predictive models.
  - Learn more…
- **PhenotypeLibrary**
  - The OHDSI Phenotype Library: a collection of community-maintained pre-defined cohorts.
  - Learn more…
- **EmpiricalCalibration**
  - Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.
  - Learn more…
- **MethodEvaluation**
  - Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods.
  - Learn more…

Cohort construction
- **EvidenceQuality**
  - Generate a wide set of diagnostics to evaluate cohort definitions against databases in the CDM.
  - Learn more…

Open Source Book
https://ohdsi.github.io/TheBookOfOhdsi/
The OHDSI ATLAS Open Source cohort designer provides a comprehensive framework for creating and managing cohort designs. The software integrates various data sources, enabling researchers to efficiently design cohorts for specific research questions. Key features include:

- **Data Sources**: Allows users to select from a variety of sources to include in their cohort definition.
- **Person**: Enables detailed analysis of demographic and health-related attributes associated with individuals in the cohort.

The cohort designer includes visual tools for exploring cohort characteristics, such as:

- **Year of Birth Distribution**: Graphically represents the distribution of cohort members by year of birth, highlighting key demographic trends.
- **Gender Distribution**: Provides a visual breakdown of the cohort by gender, showing the proportion of male and female members.
- **Race Distribution**: Illustrates the racial composition of the cohort, categorizing participants by race.
- **Ethnicity Distribution**: Offers insights into the ethnicity of cohort members, segmented into different groups.

These tools facilitate a deeper understanding of cohort composition and allow for refined cohort selection based on specific study requirements.
```
OHDSI Broadsea-Webtools Open Source

Install
- git clone ohdsi/broadsea
- docker pull ohdsi/broadsea-webtools
- docker pull ohdsi/broadsea-eunomia

Run
- docker-compose up -d

“Zero to Atlas in ten minutes”

Docker Container

App Libraries & Drivers
- Java Runtime, Javascript, JDBC database drivers

Web User Interface
- Atlas Cohort Designer

Application Services
- WebAPI Java App running in Apache Tomcat Server

1
```
OHDSI Broadsea 2.0 - FDA funded improvements

Security

- Open Source for transparency
- Docker images scanned for Common Vulnerabilities & Exposures (CVEs)
- Images cryptographically signed by OHDSI - Docker Engine can optionally verify image signature

Versioning & Meta-data

- Public Docker Hub repository retains older versions of images
- Reproducibility - e.g. a study may specify recommended Broadsea-Hades version
- Meta-data - additional provenance details included in image
- Automated build through GitHub actions (* soon)

More Broadsea Images

- Broadsea-Webtools - OHDSI Atlas
- Broadsea-Hades - RStudio Server + Hades R packages (replaces Broadsea-WebMethods)
- Broadsea-Eunomia - PostgreSQL database pre-loaded with Eunomia synthetic dataset
- Broadsea-Usagi - Perseus version of Usagi code mapping tool with web UI (* soon)
Azure Cloud Infrastructure Deployment

Microsoft Open Source Solution - Cory Stevenson & his team

- **Install Pipelines on Azure DevOps Service**
  - Clone **OHDSIOnAzure** GitHub repository
  - Set Variable Groups to customize environment

- **Create Cloud Infrastructure**
  - Execute Terraform script to create cloud infrastructure

- **Install Application Software**
  - Launch OHDSI Broadsea Docker Containers on App Service

- **Load OMOP Vocabulary Data**
  - Download OMOP vocabulary data from OHDSI Athena website and load into database

Also, “OHDSIOnAWS” open source solution (no Perseus yet) & potential future Google Cloud Platform solution.
CURE ID Pilot Results

- Pilot site: reduced ETL effort by 80% to 200 hours
- Recruiting sites to implement Perseus
  - Join VIRUS COVID-19 registry
  - Funding available
- Expanding beyond COVID
  - Meningitis
  - Sepsis
- More information:
  - CDRC@c-path.org

https://github.com/OHDSI/CureIdRegistry
Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other
Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP
Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

How to lower the ETL barrier going to OMOP with Perseus (Demonstration)
Presenter: Zachary Wang • Graduate Student, Johns Hopkins (2022 Kheiron Cohort member)

Lowering the deployment burden with the cloud
Presenter: Lee Evans • Owner, LTS Computing LLC

Distributed Machine Learning Using OMOP
Presenter: Emily Pfaff • Research Assistant Professor, University of North Carolina at Chapel Hill